2 The Basic Argument against Moral Responsibility

The best account of moral responsibility was given more than five centuries ago by a young Italian nobleman, Count Giovanni Pico della Mirandola. In his "Oration on the Dignity of Man," Pico della Mirandola explained the origins of the uniquely human miraculous capacity for moral responsibility. In the process of creation, God gave special characteristics to every realm of His great cosmos, but when His work was finished, God "longed for someone to reflect on the plan of so great a creation, to love its beauty, and to admire its magnitude," so He created humans for that role. But all the special gifts had already been bestowed on other elements of His creation, and there was nothing left for humans. So God decreed that humans "should share in common whatever properties had been peculiar to each of the other creatures"; that is, only humans would have the special power to make of themselves whatever they freely chose to be:

> The nature of all other beings is limited and constrained within the bounds of laws prescribed by Us. Thou, constrained by no limits, in accordance with thine own free will, in whose hand We have placed thee, shalt ordain for thyself the limits of thy nature. We have set thee at the world's center that thou mayest from thence more easily observe whatever is in the world. We have made thee neither of heaven nor of earth, neither mortal nor immortal, so that with freedom of choice and with honor, as though the maker and molder of thyself, thou mayest fashion thyself in whatever shape thou shalt prefer. Thou shalt have the power to degenerate into the lower forms of life, which are brutish. Thou shalt have the power, out of thy soul's judgment, to be reborn into the higher forms, which are divine. (Pico della Mirandola 1496/1948, 224–225)

This is a marvelous account of moral responsibility, which meets all the essential requirements: we make ourselves, by our own independent ab initio choices; past history, the genetic lottery, social circumstances, and cultural influences play no part. It might be hard to understand how such

special choices work—after all, *who* is doing the choosing?—but with miracles, anything is possible; besides, such miraculous events are not supposed to fall within the range of human understanding.

The delights of Pico della Mirandola's moral responsibility account notwithstanding, it does have one problem: it requires miracles. And although that was one of its charms for Pico della Mirandola and his contemporaries, it is a daunting problem for those devoted to a naturalistic world view that has no room for gods, ghosts, or miracles. The basic claim of this book is that moral responsibility belongs with the ghosts and gods and that it cannot survive in a naturalistic environment devoid of miracles. Roderick Chisholm has the right idea: "If we are responsible, and if what I have been trying to say is true, then we have a prerogative which some would attribute only to God: each of us, when we really act, is a prime mover unmoved. In doing what we do, we cause certain events to happen, and nothing and no one, except we ourselves, causes us to cause those events to happen" (1982, 32). But because—in the naturalistic system—we do not have such miracle-working powers, then (by *modus tollens*) it follows that we are not morally responsible.

Once we adopt a naturalistic world view and give up miraculous self-creating powers, it would seem an easy and obvious conclusion that we must also give up moral responsibility. But the moral responsibility system was too entrenched and the emotional underpinnings of retributive "justice" were too powerful: giving up moral responsibility was—and for many still is—unthinkable. So most philosophers pushed the argument in the opposite direction. The original argument (as Pico della Mirandola might have framed it) claimed that miraculous ultimate self-making powers are a necessary condition for moral responsibility—we do have moral responsibility; therefore, we must have miraculous self-making powers. Naturalists who reject moral responsibility agree that miraculous self-making powers are necessary for moral responsibility and conclude that because naturalism leaves no room for such powers, it thus leaves no room for moral responsibility. Those who embrace naturalism but refuse to abandon moral responsibility take a different line: we *know* that we are morally responsible, so—because miraculous self-making powers do not exist (in our natural world)—we must have been mistaken about the powers necessary for moral responsibility. Those powers were not special miraculous powers of ultimate control, but significantly more modest powers that can

fit within a naturalistic system. Thus Dennett insists that "skepticism about the very possibility of culpability arises from a misplaced reverence for an absolutist ideal: the concept of total, before-the-eyes-of-God Guilt. That fact that *that* condition is never to be met in this world should not mislead us into skepticism about the integrity of our institution of moral responsibility" (1984, 165).

Carlos Moya admonishes us not to make ultimate control a condition for moral responsibility, because doing so "is to lose sight of our ordinary practice of moral responsibility ascriptions and to look instead for an unattainable myth" (2006, 91). Later chapters examine various compatibilist efforts to lower the bar for what sort of control powers suffice for moral responsibility, but before making that examination, it is important to note the plausibility of the original—naturalistically unattainable—standard for moral responsibility.

The Case against Moral Responsibility

The traditional question is whether determinism is compatible with moral responsibility; however, the more basic issue—and the question as it is posed in most contemporary philosophical discussions—is whether moral responsibility is compatible with a naturalism devoid of miraculous powers. Bernard Williams makes it clear that it is naturalism (rather than determinism) that poses the challenge for moral responsibility:

> There may have been a time when belief in a universal determinism looked like the best reason there was of expecting strong naturalistic explanations of psychological states and happenings, but, if that was once the case, it is no longer so. It now looks a great deal more plausible and intelligible that there should be such explanations than that the universe should be a deterministic system, and it is the possibility of those explanations that itself creates the problem. (1995, 7)

There have been many arguments why moral responsibility does fit within naturalism/determinism, and those arguments are examined in subsequent chapters. But first, it is worth noting that there are powerful grounds for supposing that moral responsibility is fundamentally incompatible with naturalism. It is particularly worth noting—because the contemporary philosophical fashion is to look disdainfully at those who believe that moral responsibility is incompatible with our naturalistic world view. It is implied that anyone who understands the

sophisticated philosophical position of compatibilism will abandon the naïve notion that there is a conflict between determinism/naturalism and moral responsibility.

Arguments for compatibilism are legion and wonderful in their rich variety; the argument for the *in*compatibilism of moral responsibility and naturalism comes in several models, but it is constructed from a common foundation. The fundamental naturalistic argument against moral responsibility is that it is unfair to punish one and reward another based on their different acts, because their different behaviors are ultimately the result of causal forces they did not control, causal factors which were a matter of good or bad luck. For Lorenzo Valla, God "created the wolf fierce, the hare timid, the lion brave, the ass stupid, the dog savage, the sheep mild, so he fashioned some men hard of heart, others soft, he generated one given to evil, the other to virtue, and, further, he gave a capacity for reform to one and made another incorrigible" (1443/1948, 173).

So although some are genuinely virtuous while others are evil, and some reform their evil characters and become virtuous while others lack the resources for such reform, in all cases, the capacities for good or bad behavior are ultimately the result of their good or bad fortune, and thus there are no grounds for moral responsibility for acts or character (some may reform themselves, but that does not make them morally responsible, as the capacity for such reform is a matter of luck and not something over which they have ultimate control). Spinoza (1677/1985), Holbach (1770/1970), and Schopenhauer (1841/1960) argue that once we trace all the causes in detail, we recognize that all our acts can be traced back to earlier sources that we did not control. Thomas Nagel's problem of "moral luck" is based on the recognition that—under close scrutiny—luck swallows up the ultimate control required for moral responsibility: "If one cannot be responsible for consequences of one's acts due to factors beyond one's control, or for antecedents of one's acts that are properties of temperament not subject to one's will, or for the circumstances that pose one's moral choices, then how can one be responsible even for the stripped-down acts of will itself, if *they* are the product of antecedent circumstances outside of the will's control?" (1979, 34)

Peter van Inwagen developed his *Consequence* argument to show that moral responsibility will require a special libertarian break in the determinist/naturalist world: "If determinism is true, then our acts are the

consequences of the laws of nature and events in the remote past. But it is not up to us what went on before we were born, and neither is it up to us what the laws of nature are. Therefore, the consequences of these things (including our present acts) are not up to us" (1983, 16).

Derk Pereboom (2001, 2007) constructed his Four Cases argument to show that contrived cases in which we obviously lack adequate control for moral responsibility are (when we look closely) relevantly similar to ordinary cases in which moral responsibility is commonly assumed: in all the cases, the subjects lack the sort of control that moral responsibility requires. Galen Strawson's Basic Argument (2010) is designed to undercut the ultimate causal control necessary for moral responsibility: you do what you do because of the way you are, and so to be ultimately responsible for what you do, you must be ultimately responsible for the way you are. But you can't be ultimately responsible for the way you are (because your genetic inheritance and early experience shaped you, and ways in which you subsequently change yourself are the result of that genetic inheritance and early experience, and you are certainly not responsible for those), so you can't be ultimately responsible for what you do. All of these arguments, fascinating as they are in their detail and structure and inventiveness, are variations on a single theme. Sometimes it is presented in terms of luck; sometimes the focus is on the impossibility of making ourselves from scratch, without being limited by our raw material and our self-making skills; it may be offered in terms of basic fairness; in some versions, the focus is on the inevitability of the result given the initial capacities, but all are based on the claim that our characters (and the behavior that stems from our characters) is the product of causal forces that we ultimately *did not control*. Indeed, the argument is a naturalized version of the ancient arguments for the incompatibility of God's omnipotence with human moral responsibility, with God taking the place of nature. Some have a capacity for reform, Valla (1443/1948, 173) notes, and some don't: having or not having that capacity was not under our control.

Comparative Unfairness

I prefer to frame this fundamental challenge to moral responsibility in terms of a *comparative unfairness* argument. It is just another way of presenting the same basic argument against moral responsibility, or—more

precisely—the same basic argument to show that claims and ascriptions of moral responsibility to humans (humans who lack godlike miraculous powers of originating self-creation) are unfair. This comparative unfairness argument—which is a variation on a very old theme—goes like this. Consider two people, Karen and Louise, performing an act of moral significance: as an example, they are confronted with a situation in which their supervisor is about to make an overtly racist hiring decision, and they must object or acquiesce, knowing that a strong objection will probably block the racist decision but will also have a chilling effect on their career advancement prospects. Both Karen and Louise are intelligent persons, capable of deliberation; both are ambitious; both find racism morally repulsive; both are competent; and both are aware that the hiring decision is racist and that challenging it will be personally risky. Karen takes a courageous strong stance against this racist act, and Louise meekly acquiesces. Karen behaves in a morally upright manner, and Louise's act is morally bad. (Some may doubt that we can call an act morally good or morally bad until we know whether the actors are morally responsible; that is an issue that will be discussed later.)

Why did Karen act virtuously and Louise act vilely? There are four possibilities. One possibility is that the difference was a result of *chance*: the dice rolled, and this time Karen came up the moral winner and Louise the loser, but if we play the same case out tomorrow, the result could just as easily be the opposite. But if the result is just a matter of chance, then—as David Hume (1748/2000) so effectively argued—then neither can be morally responsible, because the events of resistance and acquiescence do not seem to belong to Karen and Louise at all: they are fortunate and unfortunate events in the world, but not the acts of Karen and Louise. Robert Kane (1985, 1996, 2002, 2007) has developed a very sophisticated argument to show that indeterminism might play some role in morally responsible acts, but he would agree that if the results were simply attributable to chance, then there is no basis for moral responsibility. A second possibility is that the different results were the product of the miracle-working powers of Karen and Louise: they make choices that are *first causes*, they act as *unmoved movers*, they originate choices through a miraculous power that transcends all natural causes and boundaries. If one offers that as an explanation of their different actions, then there is nothing more to say. With miracles, you can "explain" anything, but the price you pay is

abandoning the naturalistic-scientific framework and abandoning any hope of explaining the difference in a manner that is accessible to human inquiry and scientific investigation. The third possibility is that Karen and Louise are actually in very different *situations* (and that had their situations been reversed, their acts would also have been reversed). Thanks to the intriguing results from situationist psychological research, we now understand that seemingly insignificant differences in environmental circumstances—an admonition to hurry (Darley and Batson 1973), a lab-coated researcher urging "please continue the experiment" (Milgram 1963), finding a dime in a telephone booth (Isen and Levin 1972)—can have a profound impact on behavior: it can make the difference between stopping to help and rushing past a person in distress, and that difference (between callous disregard and kind assistance) is a significant difference indeed. And as the infamous Milgram (1963) authority experiment and the Stanford Prison Guard experiment (Haney, Banks, and Zimbardo 1973) teach us, in the right situation, most of us would perform acts of cruelty that we fervently believe we would never do under any circumstances. But if the difference between Karen's courageous act and Louise's dastardly acquiescence is the product of a difference in their situations—situations they neither made nor chose—then it is difficult to believe that they justly deserve the profoundly different treatments of reward and punishment, praise and blame. Or finally, we can insist that there is something in Karen and Louise—some strength or weakness of character—that accounts for their behavioral divergence, for if their characters and capacities—including rational capacities—were the same, and they were in identical situations, and neither chance nor miracles intervened, then Karen and Louise would perform identical acts and there would be no basis for distinguishing between their just deserts. But if we look carefully and thoroughly at the way their character traits were shaped, we recognize that ultimately they were shaped by influences and forces that were not under their control. If Karen tries harder, thinks more effectively, deliberates more thoroughly, or empathizes more deeply than Louise, then Karen's superior powers (like Louise's inferior qualities) resulted from causes far beyond her control. Or perhaps Karen worked hard to develop her own superior thinking skills, and Louise exerted no such efforts, but in that case, Karen's fortitude, as well as her strong commitment to self-improvement—qualities that certainly *do* facilitate self-improvement—ultimately can be traced back to

Karen's good developmental fortune and not to her own choices and efforts. Likewise, the inferior fortitude and commitment of Louise (which led to meager or abortive efforts toward self-improvement) were due to forces that were not under her control.

It may be tempting to say that everyone can always try harder and that therefore it is Louise's own fault that she exerted less effort toward cognitive self-improvement, so—when she does something bad because of her inferior critical thinking abilities—she justly deserves opprobrium. But that is to push this account over into the miracle-working model: it detaches *effort-making* from any causal or conditioning history, so that in the area of effort-making we are first causes or unmoved movers. When we think carefully about it, few of us imagine that our capacities to exert effort and show fortitude are under our pure volitional control: if we have great fortitude, it is because that fortitude was *shaped* and strengthened over a long and fortunate history (had we spent our younger years in circumstances in which all our efforts were failures that produced nothing of benefit—and perhaps even brought punitive responses, possibly in the form of ridicule—then we would not have the degree of fortitude we now *fortunately* enjoy: we can no more choose to exert effective sustained efforts than we can choose—at this point, with no training—to be an effective marathoner). If we refrain from appeals to miraculous powers—whether they are powers of sustained effort-making or rational deliberation—then careful comparison of the acts of Karen and Louise leaves no room to justify claims of significant differences in their just deserts.

When we look deeper and longer at exactly how their characters (including both strengths and flaws) were shaped, we find (if we renounce miraculous self-forming powers) that their characters were the product of causal forces that neither woman controlled or chose. Karen is more reflective, and perhaps more deeply committed to her nonracist values; she has a much stronger sense of positive self-efficacy:[1] self-confidence in her own ability to effectively produce positive results. Furthermore, she has a strong sense of internal locus-of-control (Rotter 1966, 1975, 1979, 1989): she believes that her *own efforts*—rather than external forces—are vitally important in shaping outcomes. All these factors are important and valuable, and they enable Karen to stand up against her racist supervisor. And Karen isn't "just lucky" to have those characteristics; she has nurtured them through her own efforts. But her capacity to nurture them and the

rudimentary powers that were there for the nurturing and further development were not there by Karen's choice and were not under her early control. Suppose Karen had recognized in herself a harmful tendency toward external locus-of-control and successfully worked to develop a stronger sense of internal control: her capacity for sustained reflection and careful self-scrutiny and her strong sense of self-efficacy to undertake self-improvement projects made such self-modification possible; those valuable resources were ultimately not of her own making or choosing, and neither those qualities nor the results that flow from them are a legitimate basis for moral responsibility. Even the example is problematic: if Karen has a strong sense of external locus-of-control (a character trait she developed at a very early age and without choice or reflection), then it is very unlikely that it will occur to her that her sense of control might be reshaped by her own powers. Now compare Karen and Louise, with the deep understanding of how their vices and virtues were shaped. Is it fair to treat Louise worse and to subject her to blame and perhaps punishment for an act she could not have avoided? Of course, if she had been a different person with different capacities and a different history, then she would have acted differently. If she had exactly the same history and resulting character as Karen, she would have acted as Karen did. In a different world, there would have been a different result, but that fact has no relevance whatsoever for the question of whether Louise justly deserves blame or punishment in the world in which she actually lives and acts and which shaped her in every detail. Louise does have flaws; does she deserve blame for them?

It is obvious that we do not make ourselves: ultimately, we are the products of an elaborate evolutionary, genetic, cultural, and conditioning history. So whether I am vile or virtuous, I am not so by my own making. It is doubtful that we can make sense of the idea of having made ourselves or chosen our own characters. Certainly, any ultimate self-making would have to occur outside the natural world: if it makes sense at all, it could only be in a world of miracles. However, some philosophers have suggested that such ab initio self-making is not required for moral responsibility and that some intermediate level of self-construction might suffice. For example, Daniel Dennett states:

> I *take* responsibility for any thing I make and then inflict upon the general public; if my soup causes food poisoning, or my automobile causes air pollution, or my robot runs amok and kills someone, I, the manufacturer, am to blame. And although

> I may manage to get my suppliers and subcontractors to share the liability somewhat, I am held responsible for releasing the product to the public with whatever flaws it has. Common wisdom has it that much the same rationale grounds personal responsibility; I have created and unleashed an agent who is myself; if its acts produce harm, the manufacturer is held responsible. I think this common wisdom is indeed wisdom. (1984, 85)

Obviously, Dennett—whose naturalist credentials are not in doubt—does not suppose we make ourselves "from scratch" in some miraculous manner. Instead, once we reach a certain level of competence, we begin to shape ourselves. But this "intermediate level" of self-making cannot support moral responsibility. If you "make yourself" more effectively than I do, it is because you have better resources for self-making; those are resources that you did not make yourself, but resources that you are lucky to have and that I am unlucky to lack. If Jan makes a better product than Kate, but Jan has the use of better raw materials, higher quality tools, and a superior work environment, then it is unfair to ascribe moral responsibility to Jan and Kate and to reward Jan while punishing Kate for their very different outputs. Perhaps at some earlier point, Jan thought carefully and chose to develop her cognitive capacities through vigorous cognitive exercise, but that added step leads back to the same destination: her beneficial cognitive exertions were the result of differences in cognitive capacities that were products of good luck and that cannot justify assertions of moral responsibility. This argument is only a preliminary sketch of one that requires much more discussion (see chapter 8), but the immediate point is that squaring moral responsibility with naturalism will not be an easy task and that there are good reasons behind the "naïve" view that naturalism/determinism is incompatible with moral responsibility.

The Unfairness Argument against Moral Responsibility

The central claim of this book is that claims and ascriptions of moral responsibility are unfair: it is fundamentally unfair to give special praise and reward to some and to blame and punish others. It is unfair because the differences in our characters and behavior are the result of causal forces that we ultimately did not choose and did not control. To examine that claim from a different angle, and to put flesh on the comparative unfairness argument sketched previously, consider the arguments of two

of the clearest and most forceful participants in the debate over moral responsibility: Galen Strawson's "regress" argument against moral responsibility and Albert Mele's critique of that argument.

Galen Strawson (1986, 28–29) formalized a well-known argument against moral responsibility: the regress argument. In its essentials, the argument goes like this. If one is to be truly responsible for how one acts, one must be truly responsible for how one is, morally speaking. To be truly responsible for how one is, one must have chosen to be the way one is. But one cannot really be said to choose (in a conscious, reasoned fashion) the way one is unless one already has some principles of choice (preferences, values, ideals) in the light of which one chooses how to be. But then to be truly responsible on account of having chosen to be the way one is, one must be truly responsible for one's having *those* principles of choice, but then one must have chosen them, in a reasoned, conscious fashion. But that requires that one have principles of choice. And thus the regress.

Alfred Mele develops a powerful critique of Strawson's regress argument, focusing on a vital premise of the argument: to be truly responsible for how one is, one must have *chosen* to be that way. He offers a charming example to support his critique: the case of Betty, a six-year-old child with a fear of the basement. Betty knows that no harm has come to herself or others when they have ventured into the basement, and she recognizes that her older sister has no fear of the basement. Betty decides that her fear is "babyish," and that she will take steps to overcome it. Her plan is simple but effective: she will make periodic visits to the basement until she no longer feels afraid there. As Mele states, "If Betty succeeds in eliminating her fear in this way, this is an instance of intentional self-modification" (1995, 223).

Clearly such cases of "intentional self-modification" are possible; as Mele argues, there is no reason to suppose that they stem from "an infinitely regressive series of choices":

> Betty's choice or decision to try to eliminate her fear need not rest on any attitude that she *chooses* to have. Desires and beliefs of hers might ground her choice—and her judgment that it would be best to try to eliminate the fear—without her having chosen to have any of those desires (or beliefs). Can she nevertheless be "truly responsible" for her choice and her behavior? If it is claimed that *true responsibility* for any choice, *by definition*, requires that the agent have chosen "in a conscious,

reasoned fashion" an attitude that grounds the choice, it is being claimed, in effect, that the very definition of 'true responsibility' entails that possessing such responsibility for any choice requires having made an infinitely regressive series of choices. (1995, 223–224)

But Mele insists that we "should want to have nothing to do with *this* notion of responsibility, nor with any corresponding notion of free action" (1995, 224). So what sort of freedom *should* we find desirable? Mele's answer is as follows:

In ordinary practice (at least as a first approximation), when we are confident that self-reflective, planning agents have acted intentionally, we take them to have acted freely *unless* we have contrary evidence—evidence of brainwashing, compulsion, coercion, insanity, or relevant deception, for example.

In the same vein, we take Betty to have freely tried to eliminate her fear. Our learning that she did not choose to have any of the attitudes on the basis of which she chose or decided to make the attempt will not incline us to withdraw the attribution of freedom, unless we are inclined to hold that free action derives from choices made partly on the basis of chosen attitudes or, at least, that any action etiology that includes no such choice is a freedom-blocking etiology. Those who have this latter inclination are, I suggest, in the grip of a crude picture of the freedom of an agent with respect to an action (or "practical freedom," for short) as a *transmitted* property—a property transmitted from above by earlier free behavior, including, of course, choice-making behavior. It is impossible for such a picture of practical freedom to capture the freedom that it is designed to represent, for reasons that Strawson makes clear: the picture requires an impossible psychological regress. And it ought to be rejected. Practical freedom, if it is a possible property of human beings, is, rather, an "emergent" property. It must be, if some of us are free agents (i.e., agents who act freely, in a broad sense of 'act' that includes such mental actions as choosing) and none of us started out that way. (1995, 224–225)

This passage is a superb description of how the capacity for "practical freedom" develops. The development of that capacity requires nothing mystical, and it does not result in a vicious regress. As Mele makes clear, most of us do develop that capacity and become free agents, and "none of us started out that way."

Mele gives a marvelous account of "free agency/practical freedom" and its emergence. But its virtues notwithstanding, the account does nothing to establish moral responsibility; to the contrary, careful examination of Mele's excellent account of freedom and its enriched development soon undercuts any claims of moral responsibility. Place alongside Betty her six-year-old twin brother, Benji, who also suffers from fear of his basement

(and who, like Betty, knows that no harm has befallen those who venture there). Benji also regards this fear as "childish" and wishes to get beyond it. But Benji is a little—just a little—less self-confident than his sister. Rather than taking bold steps to deal with his fear, Benji decides to wait it out: maybe I'll grow a bit bolder as I grow older, Benji thinks; besides, Mom is plenty strong and courageous, so there's no need for me to make an effort that might well fail. Betty has thought up a good plan, Benji recognizes, but well-planned projects often come to a bad end, like that well-thought-out plan to stand on a chair to reach the cookie jar. Benji is not quite as strong as his sister, in some very significant respects. He does not have her high level of self-confidence (or sense of self-efficacy); his sister has a strong internal locus-of-control, but Benji is inclined to see the locus-of-control residing in powerful others. And although Betty is well on her way to becoming a chronic cognizer (to be discussed shortly), Benji has developed significant tendencies toward cognitive miserliness (the abysmal failure of that well-thought-out campaign to liberate the cookie jar left a deep mark); that is, even at this tender age, Betty and Benji already have significant differences (not of their own making or choosing) in what psychologists call "need for cognition" (Cohen, Stotland, and Wolfe 1955).

Cognitive misers (Cacioppo and Petty 1982; Cacioppo et al. 1996) do not enjoy thinking—especially careful in-depth abstract thinking—and they tend to make decisions more quickly, with less deliberation, and with less attention to all the significant details; in contrast, *chronic cognizers* take pleasure in thinking, eagerly engage in careful extended deliberation, and reflect in more detail and at greater depth before making decisions. Like their differences in self-efficacy and locus-of-control, this early difference in need for cognition is likely to have profound effects: those with a weaker need for cognition (the cognitive misers) are more likely to be dogmatic and closed-minded (Cacioppo and Petty 1982; Fletcher et al. 1986; Petty and Jarvis 1996; Webster and Kruglanski 1994), and they are more likely to avoid or distort new information that conflicts with their settled beliefs (Venkatraman et al. 1990). In contrast, those who are fortunate enough to be shaped as chronic cognizers tend to have greater cognitive fortitude (Osberg 1987), stronger curiosity (Olson, Camp, and Fuller 1984), be more open to new experiences and new information and more careful in evaluating new information (Venkatraman et al. 1990; Venkatraman and Price 1990; Berzonsky and Sullivan 1992), and more successful in solving

complex problems (Nair and Ramnarayan 2000, 305). That's not to say that Benji is doomed to a terrible fate, or that he will never reach a level of competence, or that he will never be capable of making his own decisions (and, overall, benefit by doing so). But it is to say that Benji's incipient "relatively sophisticated intentional behavior" will have fewer resources to draw upon than those available to his sister.

One of the best parts of Mele's argument is his very plausible account of the long-term results from six-year-old Betty's intentional self-modification:

> Agents' free choices and actions have significant psychological consequences for them. By choosing and acting as we do, we affect our psychological constitution—sometimes, even, *intentionally* affect it, as young Betty did. Further, successes like Betty's may have important consequences for agents' psychological constitutions well beyond the immediate present. Betty's success in conquering her fear may, for example, enhance her self-esteem, expand her conception of the range of things she can control, and contribute to her deciding to try to conquer other fears of hers. Her successful effort at self-modification regarding her fear of her basement may lead to bigger and better things in the sphere of self-modification as a partial consequence of its relatively proximal effects on her psychological condition; and given that the effort was freely made, a *free* action of Betty's will have contributed to the psychological changes. Of course, the more proximal bigger and better things may lead to more remote ones that are bigger and better yet. Seemingly minor successes at self-modification may have, over time, a major impact on one's character. (1995, 229)

Both Betty and Benji make free choices, and those choices have significant impact on their formed characters and their subsequent choices stemming from those qualities of character. Mature Betty has a strong internal locus-of-control, believing that she herself has significant control over the most important events in her life; Benji believes that much of what happens to him is outside his power to control, being in the hands of powerful others (perhaps God). Betty has a powerful sense of confident self-efficacy for most of the projects that are important to her: she believes that she is very good at acting and controlling (including acting to change herself, should she find faults that need changing). Benji's sense of self-efficacy is substantially weaker: he is not very confident that he can carry out his valued projects successfully (including any self-improvement projects). Mature Benji wants to stop smoking, and he might try to do so—but he doesn't really believe that he has the resources to succeed (and because one of the

needed resources is a strong sense of self-efficacy, he is probably correct in expecting failure). Betty's strong resources give her a very generous measure of freedom. Benji also chooses and acts freely, though without the rich freedom resources enjoyed by Betty. And as Mele makes clear, those differing resources—though to a significant degree self-made—were shaped by initial resources "that she did not choose to have."

The initial resources, and the early choices that stem from them, are of great importance. As Mele notes:

> One's earliest or most primitive free choices are not themselves made on the basis of freely chosen attitudes. It cannot be otherwise; the earliest free choices of an agent cannot themselves be made, even partly, on the basis of *other* free choices of the agent. But this does not preclude one's developing into a person like Betty: a self-conscious, self-reflective, self-assessing agent who can intentionally and freely undertake to eliminate or foster an attitude in herself—and succeed. Success in such endeavors can have consequences for the agent's developing character. The same is true of *failure*. (1995, 230)

So Betty, with her somewhat stronger resources, attempts to eliminate her basement fear and succeeds. Benji takes a more passive path in dealing with his "babyish" fear, and his lesser efforts are a failure. And as Mele notes, both the success and failure "can have consequences for the agent's developing character" (1995, 230), as Betty waxes in self-confidence and cognitive fortitude, while Benji wanes. Twenty years pass, as both continue their divergent development paths. Both have grown up as members of the privileged race in a profoundly racist society, and both have enculturated the racist values of their society, and both have employed their powers of self-assessment to question those racist values. Both Betty and Benji wish to change, but Betty—due to her early success—has more resources and is more successful. She really does change, and she changes because she chooses to do so, and because she exerts the effort and the intelligent planning to succeed. Benji does not, because he is more acquiescent, or less self-confident, or less reflective, or less self-assessing; in short, because his tools for further self-development are not as good. Betty has better resources, and she uses them more effectively, but whether she deserves credit and Benji deserves blame is a very different matter.

Benji does have some powers of reflection and self-assessment—not powers as robust as his sister's, but to a lesser degree. But Benji's reflective self-assessment may result in his becoming resigned to and even contented

with his lot: like Eliot's J. Alfred Prufrock, Benji concludes that "I am not Prince Hamlet, nor was meant to be," and I'm not really equipped to undertake major changes or challenges: "I know I'm a racist, and I know that's not good. But so are my friends, and that's who I am, and change is very hard for me. Besides, I'm not very good at giving up bad habits: look at my failed attempts to stop smoking. Changing racism is beyond my powers, anyway; it will have to be done by our leaders. Better not to think too hard about it." Betty is now a civil rights campaigner and is morally very good; Benji is a racist who acquiesces in the racist status quo and is morally bad. Both make free choices (though Betty's are freer than Benji's). Both want and need the freedom to make their own choices (Benji cannot make choices and carry them through as effectively as Betty does; that doesn't mean that he does not wish to make his own choices, and he would deeply resent Betty trying to run his life for him). Both can and do exercise *take-charge* responsibility (the responsibility—distinguished from moral responsibility—for making one's own decisions concerning one's life) as discussed in greater detail in chapter 6; though again, Betty—with her stronger sense of self-efficacy and greater cognitive fortitude—exercises it much better than does Benji. They do act freely, but that freedom does not establish moral responsibility and just deserts. Betty questions and challenges the system, and Benji acquiesces (as noted earlier, Benji's tendency toward cognitive miserliness leads him to avoid new information that would upset his settled beliefs and require him to think carefully about his views). This difference may be a very serious one indeed, if they are both growing up in a viciously racist society. Betty really is a stronger person, indeed a better person, but whether she deserves credit for her better character (much of it self-formed) and her superior behavior is a different question altogether.

Mele offers a clear and valuable reminder of something Aristotle (350 BC/1925) emphasized long ago: our choices today shape our choices and our characters of tomorrow. If you want to be a person of integrity tomorrow, then do not lie and cheat today. Studying the history that shaped us and how we emerged as free actors making our own choices is very important, and that includes the study of the critically important *initial capacities* in our earliest choices and how those capacities are fostered or inhibited. But studying that history with the hope of finding grounds for moral responsibility is a futile hope.

Timothy O'Connor recognizes the important influence of our early unchosen reflective powers and propensities, but he seems to think we can "grow out of" those influences:

> We come into the world with powerful tendencies that are refined by the particular circumstances in which we develop. All of these facts are for us merely "given." They determine what choices we have to make and which options we will consider (and how seriously) as we arrive at a more reflective age. However, presuming that we are fortunate enough not to be impacted by traumatic events that will forever limit what is psychologically possible for us, and, on the positive side, that we are exposed to a suitably rich form of horizon-expanding opportunities, the structure of our choices increasingly reflects our own prior choices. In this way, our freedom *grows* over time. (2005, 219–220)

Our characters and our subsequent choices do reflect "our own prior choices"; as we develop, our characters thus become more our own. But whether our own characters are good or bad, strong or weak, we are not morally responsible; unless our choices can miraculously transcend our causal history, our characters and subsequent choices are shaped by given backgrounds that set the direction of further development. Our freedom may grow (along with a stronger sense of self-efficacy and internal locus-of-control and cognitive fortitude), as in the case of Betty, or we may become less confident and more rigid and less reflective, as does Benji. When we look carefully at the differences in their developed characters, we recognize—unless we trust in miracles (such as a power of reason to transcend causal histories), or attribute the difference to chance events they did not control—that those differences were the product of early differences in capacities or in circumstances which they did not control and for which they are not morally responsible. And absent such ultimate control, it is unfair to reward one and punish the other, or praise one and blame the other; that is, it is unfair to treat them in dramatically different ways. Their characters and behavior are their own, but that does not make Betty and Benji morally responsible.

Kane's Argument for Ultimate Responsibility

Robert Kane has made a remarkably innovative and thorough effort to establish room for moral responsibility within a thoroughly naturalistic world. Kane faces the challenge squarely: he refuses to take refuge in

attempts to block examination of how our characters were shaped and the differences in our histories, and he rejects facile notions that we can somehow make ourselves without regard to the self-making capacities from which we start. To the contrary, Kane insists that moral responsibility requires genuine *ultimate control*: the "before-the-eyes-of-God" ultimate control that many defenders of moral responsibility dismiss as too strong a requirement. No one has confronted more directly, or struggled more vigorously, with the problem of justifying moral responsibility without compromising naturalism. Kane's remarkable efforts to establish ultimate grounds for moral responsibility without miracles or mysteries are worthy of examination on their own merits; examination of his strong and straightforward arguments offer a clear setting in which to bring the basic argument against moral responsibility—the comparative unfairness argument, as I have framed it—into clearer focus.

Kane attempts to establish naturalistic ultimate control by incorporating a crucial element of indeterminism (he insists it is not chance) into his impressive account of crucial self-forming acts. No brief discussion can do justice to the subtlety and sophistication of Kane's libertarian theory, but the crux of his position is this: our freedom and moral responsibility require the existence of "self-forming acts," in which we genuinely will both of two different open alternatives that cannot both be fulfilled; in the course of this incompatible willing, our neural networks create the right conditions for a genuine indeterminism (in which the random movement of a subatomic particle is amplified by the chaos created by conflicts of neural networks) such that either of these willed events can occur, but whichever event actually occurs, it is an act that we willed, an act for which we have reasons (reasons that we endorse), an act that is not coerced, an act that we acknowledge as our own and for which we take responsibility, an act which results from our own effort of will; that is, for both the genuinely possible acts, we have *dual-control* responsibility (Kane 2002).

In applying a comparative unfairness critique to this model, consider (instead of Betty and Benji) Betty and Barbara: Barbara is identical in all relevant respects to Betty (identical in levels of need for cognition, cognitive abilities, sense of self-efficacy, locus-of-control, rational and empathetic capacities), and they confront identical situations. Both Betty and Barbara are striving to overcome their developed racist characters, and they are also striving to hang onto their comfortable racist beliefs that are

endorsed by their friends and community. At a crucial point, the result is indeterminate: Betty and Barbara are identical persons exerting identical efforts, an element of genuine indeterminism enters the equation, and Betty chooses to reject racism while Barbara chooses to remain a racist (of course Barbara will not describe the result as "remaining a racist"; she might describe it as "preserving cultural heritage"). In both cases, the choices are their own (and Kane does a superb job of making a case for dual ownership of either act), but is it really fair to blame one and praise the other? Certainly, one is now good and the other bad, but do they justly deserve differences in treatment for their different character traits? Remember, the difference in outcome is not due to differences in cognitive fortitude or curiosity or openness to new ideas or sense of self-reliance (all of which can be traced back to causes for which Benji and Betty and Barbara clearly are not morally responsible); rather, the difference must stem from indeterminism—ultimately, in Kane's model, to the amplified motion of a subatomic particle. Thus the difference between Betty and Barbara—which is now profound—is not the result of their control, but is the result of an indeterminate random roll of a subatomic particle. Both Betty and Barbara can rightly acknowledge their resulting characters as their own, but the question is not whether their characters and acts are their own, but whether they are morally responsible for them.

Barbara remains a racist, and Betty has renounced racism; both Barbara and Betty now endorse (Kane 2007, 33) those views (both of them were willing the result they now endorse while also willing the opposite result). And both are happy to "take responsibility" (41) for their resulting different characters. All of these factors contribute to an important and psychologically healthy sense of ownership of one's own character and control over what one does and what one becomes (the benefits of that sense of control is discussed in subsequent chapters, particularly chapter 6). The sense of control is not only healthy, but also legitimate: Betty and Barbara (and to a lesser but important extent, also Benji) really do exercise important control over their choices and development. But when we focus in on whether they have the ultimate control (that Kane acknowledges as essential for moral responsibility), we face a very different question. What is the difference between Betty and Barbara, for which the former deserves praise and the latter blame? The difference is that one rejects and the other embraces racism, which is a very significant difference that is likely to lead

to even more significant differences (for example, Barbara is likely to become more dogmatic and closed-minded as she struggles to preserve her racist beliefs in the face of countervailing evidence). But although there is much that Betty and Barbara did control, they did not—in Kane's indeterminist scenario—control the development of that difference. The difference between otherwise identical Barbara and Betty at that crucial indeterminate moment is that a subatomic particle bounced one way in Betty and a different way in Barbara, which is what resulted in their now different characters (different characters they both endorse and with which they identify); that key difference is not one they ultimately control, and not one for which they legitimately can be blamed or praised.

Drawing Conclusions Concerning Betty and Benji

When we encounter Betty and Benji, we have a number of possibilities. First, we can willfully ignore the detailed differences in their capacities, insist that everyone is equal on the same plateau and thus is morally responsible: there are no relevant differences among us (an argument that will be critically examined in chapter 12). Second, we could argue that once Betty and Benji have emerged to that level (regardless of how they did it), they have special rational powers that transcend all differences in details, and they can go in any direction and develop any capacities of unlimited strength. But naturalists know too much about how we are shaped—and the psychological factors affecting our rational powers—to draw any such conclusion concerning godlike powers of reason. Third, we can insist that we do have magical self-construction initial powers, as the existentialists claim: that somehow we choose ourselves, or make ourselves ab initio, that we are self-caused in some absolute (nonnatural) manner. Fourth, we could reject mysterious initial self-creation, but insist—with C. A. Campbell (1957)—that along the way we have the special miraculous power of making choices using powers of special contracausal free will that cancel out the effects of our differing initial conditions. Fifth, we can admit that we don't have such magical powers of initial or intermediate self-creation, but claim that the initial starts were generally fair and so the results are fair: as Dennett (1984) attempts for roughly equal starts, and Sher (1987) proposes for overall equal talents (a line of argument critiqued in chapter 7). Sixth, we can attempt to find space for moral responsibility

in special instances of indeterminism (Kane's model). Or finally, we can look carefully at how we were shaped and the differences in our starting abilities and in our situations, reject mysteries and miracles, and deny moral responsibility.

The story of Betty and Benji requires no nefarious neurosurgeons and no peremptory puppeteers—or any kind of devious or coercive intervener. Mele's story of Betty is a mundane, plausible, psychologically sound account of the development of Betty's capacity for practical freedom; the account of Benji's development has the same features. The accounts do not require science fiction; they require only that we look closely at the details of how our capacities are formed (including the full process of self-formation), and at how differences in those capacities result. If we start with differences, then (barring differences in racing luck, or a positive intervener—the proverbial kindly priest or concerned coach—or some other factor for which we are not responsible) we end differently. That conclusion doesn't mean that we are not largely self-made, or that we cannot exercise effective choice, but it does mean that we are not morally responsible. Betty isn't just lucky to be so strong and virtuous—after all, much of her strength and ability was shaped by her own successful efforts. But Betty (as compared to Benji) is lucky to have had the start that enabled her to become the person she is, and Benji is unlucky to have had fewer initial developmental powers. The question is not whether Betty can develop so, employing her own developing abilities; she can, as Mele insightfully describes. Nor is it a question of whether Betty can accomplish much (she can) or whether her accomplishments flow from her own strong and resourceful character (they do). The question is whether she deserves special credit and Benji deserves special blame (they don't). Once we recognize that freedom can be distinguished from moral responsibility, and that having good (or bad) qualities of character can be distinguished from being morally responsible for those character qualities, then it is clear that our best account of the development of freedom is not an account that justifies claims and ascriptions of moral responsibility.

Benji is somewhat free (he certainly should not be denied the opportunity to make his own decisions); Betty is much freer. But neither can claim ultimate responsibility; unless they have ultimate, "before-the-eyes-of-God" moral responsibility, then moral responsibility is unfair. It may also be "unfair" that Benji starts with less capacity for free will than Betty does;

it's also "unfair" that some are born sound and others with severe disabilities. Those are differences that we wish to mitigate (not by handicapping the advantaged, obviously, but by improving the opportunities for the disadvantaged), but there is no question of blaming/praising for those critical initial differences. Life is not fair, true enough, but just deserts must be fair, and the natural lottery of genetic traits and early conditioning is not a fair manner of distributing just deserts. Just deserts and moral responsibility require a godlike power—the existentialist power of choosing ourselves, the godlike power of making ourselves from scratch, the divine capacity to be an uncaused cause—that we do not have. Moral responsibility is an atavistic relic of a belief system we (as naturalists) have rejected, for good reason. Freedom—and its enhancement—fits comfortably with our natural world and our scientific understanding of it; moral responsibility does not.

The basic problem for any naturalistic defense of moral responsibility is that we are each different in our capacities and talents and cognitive abilities and fortitude; careful comparisons of those differences in character and history soon undercut any claims or ascriptions of moral responsibility. Those differences make it unfair to blame one and reward another for their differences in behavior. On the naturalist—nonmiraculous—view, if there is a difference in behavior, then there must be a difference in circumstances, influences, or abilities. This view is not intended as a conclusive argument against moral responsibility. But it should serve to establish that the burden of proof rests on those who claim that moral responsibility is compatible with naturalism. Absent such proofs, it is difficult to see how moral responsibility can fit within the naturalistic worldview. Furthermore, as Richard Double has argued (2002) the burden of proof falls heavily on those who claim there is moral responsibility, because they are proposing that we blame and punish people for their misdeeds, and justifying such painful special treatment requires a very strong proof that it is being imposed fairly. This book is an effort to show that—as naturalists—we should reject all claims and ascriptions of moral responsibility. The moral responsibility system has long since outlived the very limited advantages it offered, and it should be replaced—in law, government, education, philosophy, and common belief—by a system that will greatly reduce both physical and psychological harm and will open paths to individual and social progress.

The goal of this book is to show that moral responsibility cannot be justified, that the major arguments in support of moral responsibility fail, that the moral responsibility system is severely flawed, and that the world would be better if belief in moral responsibility vanished from the Earth. But I am not claiming that development and further refinement of the moral responsibility system was altogether bad. To the contrary, the initial development of the moral responsibility system was beneficial: it certainly is an improvement over the more primitive impulse to simply strike back (whoever slays a man shall be slain): we must have justification for striking back, and with the justification comes a wide range of exceptions and exemptions that lessened the extent of harmful punishment. Our system of retributive justice (though I believe it has outlived its early usefulness) was an enormous step forward from lynch mobs and personal vendettas.

Furthermore, I am certainly not suggesting that the rich and fascinating range of arguments in support of moral responsibility have been useless. Though I believe they fail to support moral responsibility, they have provided important insights into questions of personal identity, ethics, free will, and many other areas. Daniel Dennett and John Martin Fischer fail (I claim) to establish grounds for moral responsibility, but in the course of their efforts they have drawn a much clearer picture of the many important varieties of control, their value, the distinctions among them, and their enormous psychological significance. Harry Frankfurt and Gerald Dworkin have not established grounds for moral responsibility, but they have developed a vitally important account of the deeper psychological levels of human desire and will and have therefore greatly improved our understanding of human freedom and constraint. If Alfred Mele's work does not justify moral responsibility, it loses none of its subtle insights into the complex development of human character. And even if Robert Kane's extraordinary model of ultimate self-forming acts fails to support moral responsibility, it is a remarkably clear and honest guide to the conditions required for genuine moral responsibility.

Finally, although some people insist on moral responsibility as a means of justifying greed and exploiting desires for vengeance (some politicians spring to mind), I do not believe that those are the motives of most of the philosophical defenders of moral responsibility. Though the motives for defense of moral responsibility have been many, some of the most dedicated proponents of moral responsibility are certainly not motivated by

greed and vengeance; instead, many of them—such as Dostoyevsky (1864/1961), William James (1890), and William Barrett (1958)—want to protect a power of special creativity: the power to be the genuine author, the original source, of something distinctively new—the desire to refute Solomon's depressing insistence that "there is nothing new under the Sun." Robert Kane, though he wants moral responsibility, wants it for much more than a justification of striking back or claiming reward; he wants to be a genuine starting point, an *originator* who is more than a link in a deterministic chain (Kane 1985, 177–178). This issue is examined in chapter 14, but the immediate point is this: at least some of those who have struggled to support a workable account of moral responsibility in the face of scientific challenge have been motivated by goals far more attractive than vengeance and greed.

The case against moral responsibility is a powerful one, on both moral and pragmatic grounds. As the scientific understanding of human behavior expands, the case against moral responsibility grows stronger, while serious flaws are exposed in the arguments supporting moral responsibility. Substantiating those claims is the task of the remainder of this book. But before going into that work, it is essential to examine free will. There is no plausible naturalistic account of free will that can support the weight of moral responsibility: that is the focus of the chapters after the following chapter. The examination of free will in the following chapter makes three claims: the traditional close linkage of moral responsibility to free will is a mistake; the effort to concoct an account of free will that can bear the burden of moral responsibility has resulted in a severely deformed account of free will; and there is a naturalist account of free will that is more empirically plausible, does not support moral responsibility, and can flourish in the absence of moral responsibility.